DOCUMENT RESUME

ED 144 949 TM 006 230

AUTHOR Dalton, Starrett
TITLE [Shrinkage In the Squared Multiple Correlation Coefficient and Unbiased Estimates of Treatment Effects Using Omega Squared.]
PUB DATE [Apr 77]
NOTE 10p.; Paper presented at the Annual Meeting of the American Educational Research Association (61st, New York, New York, April 4-8, 1977)

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS *Analysis of Variance; *Multiple Regression Analysis; Sampling; *Statistical Analysis; Statistical Bias; *Statistical Studies

ABSTRACT

The amount of variance accounted for by treatment can be estimated with omega squared or with the squared multiple correlation coefficient. Monte Carlo methods were employed to compare omega squared, the squared multiple correlation coefficient, and the squared multiple correlation coefficient to which a shrinkage formula had been applied, in terms of bias and precision. The squared multiple correlation coefficient to which the shrinkage formula had been applied and omega squared produced estimates which were negligibly biased. The bias in the squared multiple correlation coefficient, while consistently positive, decreased as the sample size increased and was too small to be of practical importance when the sample size was 50 or larger. Omega squared, the squared multiple correlation coefficient and the squared multiple correlation coefficient to which the shrinkage formula had been applied were all most precise with large samples and least precise when treatment effects were moderate in magnitude. (Author/MV)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED144949

U S DEPARTMENT OF HEALTH
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Shrinkage in R^2 and Unbiased Estimates of Treatment Effects Using $\hat{\omega}^2$

Starrett Dalton

University of California, Riverside

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

STARRETT DALTON

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

TM006 230

Running head: Shrinkage in R^2 and Unbiased Estimates

Abstract

The amount of variance accounted for by treatment can be estimated with $\hat{\omega}^2$ or with R^2 (symbolized as R_c^2 after a shrinkage formula has been applied). Monte Carlo methods were employed to compare $\hat{\omega}^2$, R_c^2, and R^2 in terms of bias and precision. R_c^2 and $\hat{\omega}^2$ produced estimates which were negligibly biased. The bias in R^2, while consistently positive, decreased as sample size increased and was too small to be of practical importance when $n \geq 50$. $\hat{\omega}^2$, R_c^2 and R^2 were all most precise with large samples and least precise when treatment effects were moderate in magnitude.

Shrinkage in R^2 and Unbiased Estimates of Treatment Effects Using $\hat{\omega}^2$

William Hays (1963) encouraged investigators to provide estimates of the magnitude of treatment effects in addition to the statistics and probability levels which are customarily reported in analysis of variance. He offered omega squared ($\hat{\omega}^2$) as an unbiased estimator of the parameter eta squared (η^2).

$$\hat{\omega}^2 = \frac{SS_B - (J-1)MS_\omega}{SS_T + MS_\omega}$$

SS_T = sum of squares total

SS_B = sum of squares between groups

MS_w = mean of squares within groups

J = number of groups

and

$$\eta^2 = \frac{\sigma_y^2 - \sigma_\varepsilon^2}{\sigma_y^2}$$

σ_y^2 = variance of the dependent variable

σ_ε^2 = the common homogeneous variance of the Y_{ij} about y_j

η^2 = a population measure of the magnitude of treatment effects relative to the total variance in the experiment

Carroll and Nordholm (1975) have demonstrated that another estimate of η^2 (epsilon squared) developed earlier by Kelley (1935)

is as accurate as ω^2. Nevertheless ω^2 appears to be most popular among researchers who use analysis of variance.

Researchers who use regression analysis for testing the same hypotheses tested in analysis of variance (e.g., McNeil, Kelly, and McNeil, 1975; Namboodiri, Carter and Blalock, 1975) use the squared multiple correlation coefficient (R^2) to estimate the proportion of variance accounted for by treatment (i.e., η^2). Because R^2 is positively biased a shrinkage formula is commonly used to provide a value of R^2 corrected for bias (R_c^2).

$$R_c^2 = 1 - (1-R^2)\left(\frac{N-1}{N-k-1}\right)$$

R_c^2 = R^2 after correction for bias

N = number of subjects

k = number of independent variables

In the past, the selection of R_c^2 or $\hat{\omega}^2$ to estimate η^2 has been determined by the researcher's preference for traditional analysis of variance or regression analysis. Because no one has compared these statistics in terms of bias and precision, no evidence is available for selection on the basis of accuracy. The primary purpose of this study, therefore, is to compare the sampling distributions of R^2, R_c^2, and $\hat{\omega}^2$ in terms of bias and precision.

Procedure

Using Monte Carlo methods, the study was conducted in the context of a three level, one way analysis variance. It was assumed the results would generalize directly to higher order analysis of variance. The values of η^2 selected were .00, .05, .15, .40. .75, and .90. Sample sizes ranged from N=30 (10,10,10) to N=600 (200, 200,200). Error terms were randomly selected from normal populations with homogeneous variances. Carroll and Nordholm (1975) studied the effect of violating the assumptions of normality and homogeneity when estimating η^2 using $\hat{\omega}^2$. The reader interested in this problem is referred to their work. For each value of η^2 and for each sample size 1000 simulated experiments were conducted. Data were generated by the pseudorandom number generator in the IBM scientific package.

Results

Bias. For each value of η^2 and for each sample size the mean and standard error (SE) were calculated for R^2, R_c^2, and $\hat{\omega}^2$ (Table 1). R_c^2 and $\hat{\omega}^2$ produced estimates of η^2 which were negligibly

---

Insert Table 1 about here

---

biased. The bias in R^2, while consistently positive, decreased as sample size increased and was too small to be of practical importance

when $n \geq 50$.

Precision. As reflected in the standard errors, the imprecision of all three statistics decreased as sample size increased. Imprecision was greatest in the middle of the range of η^2 (i.e., for η^2=.15 and η^2 =.40). For small sample sizes R^2 was very slightly, yet consistently more precise than R_c^2 and $\hat{\omega}^2$.

Discussion

Any attempt to pinpoint a tendency in R_c^2 or $\hat{\omega}^2$ to be positively or negatively biased, based upon the present data, would probably be misleading. The mean values of R_c^2 and $\hat{\omega}^2$ always fell within two SEs of the population value (η^2) and 92% of the time both estimates were within one standard error of η^2. The mean value of $\hat{\omega}^2$ never differed from the population value by more than one SE. By these standards $\hat{\omega}^2$ is slightly less biased than R_c^2, however, this conclusion is overshadowed by the tendency of both measures to yield accurate estimates in terms of bias.

When Hays developed $\hat{\omega}^2$ he stressed the importance of estimating the magnitude of treatment effects when sample sizes were large. "Virtually any study can be made to show significant results if one uses enough subjects, regardless of how nonsensical the content may be." It is indeed ironic that the use of $\hat{\omega}^2$ (or

R_c^2) instead of the familiar R^2 is least necessary when sample size is large. In this case the estimates of experimental effects using R^2 show only negligible bias. In the present study the mean value of R^2 ($n \geq 50$) never differed from η^2 by more than one standard error. This degree of bias should not concern most applied researchers. In addition, R^2 is routinely produced by many packaged computer programs and consequently is more readily available than R_c^2 and $\hat{\omega}^2$.

In conclusion, it appears that R_c^2 and $\hat{\omega}^2$ may serve equally well in estimating the magnitude of experimental effects. Both are superior to R^2 in this regard when $n \leq 30$. For large samples R^2 shows little bias and in many circumstances may be sufficient as an estimate of $\hat{\omega}^2$.

References

Carroll, R. M. & Nordholm, L. A. Sampling characteristics of Kelley's and Hays' . Educational and Psychological Measurement, 1975, 35, 541-554.

Hays, W. L. Statistics for Psychologists. New York: Holt, Rinehart and Winston, 1963.

Kelly, T. L. An unbiased correlation ratio measure. Proceedings of the National Academy of Sciences, 1935, 21, 554-559.

McNeil, K., Kelly, F., & McNeil, J. Testing Research Hypotheses Using Multiple Linear Regressions. Carbondale: Southern Illinois University Press, 1975.

Namboodiri, N. K., Carter, L. F., & Blalock, H. M., Jr. Applied Multivariate Analysis and Experimental Design. New York: McGraw Hill, 1975.

Table 1

Means and Standard Errors (SE) of the Multiple Correlation Coefficient (R^2), Omega Squared (ω^2), and the Corrected Multiple Correlation Coefficient (R_c^2) for Each Condition

Sample size	Estimates of η^2	Eta Squared (η^2) .00 X	.00 SE	.05 X	.05 SE	.15 X	.15 SE	.40 X	.40 SE	.75 X	.75 SE	.90 X	.90 SE
	R^2	.069	.065	.116	.093	.209	.117	.443	.118	.771	.059	.909	.024
10	ω^2	.000	.068	.049	.097	.147	.123	.394	.125	.748	.064	.899	.027
	R_c^2	.000	.070	.051	.099	.150	.126	.402	.126	.754	.063	.902	.026
	R^2	.023	.022	.070	.048	.167	.068	.411	.073	.755	.037	.903	.015
30	ω^2	.000	.023	.049	.049	.146	.069	.395	.075	.748	.038	.899	.015
	R_c^2	.000	.023	.049	.049	.148	.070	.397	.075	.750	.037	.900	.015
	R^2	.013	.014	.063	.035	.161	.051	.408	.054	.753	.027	.901	.011
50	ω^2	.000	.014	.050	.036	.149	.051	.398	.055	.749	.027	.899	.011
	R_c^2	.029	.014	.023	.037	.126	.053	.382	.056	.743	.028	.897	.011
	R^2	.007	.007	.057	.024	.157	.036	.405	.039	.753	.020	.901	.008
100	ω^2	.000	.007	.051	.024	.151	.036	.401	.039	.750	.020	.900	.008
	R_c^2	.014	.007	.038	.025	.139	.037	.393	.040	.748	.020	.899	.008
	R^2	.004	.004	.055	.020	.154	.030	.404	.032	.752	.016	.901	.007
150	ω^2	.000	.004	.050	.020	.150	.030	.401	.032	.751	.016	.900	.007
	R_c^2	.000	.004	.050	.020	.151	.030	.401	.032	.751	.016	.901	.007
	R^2	.003	.003	.054	.017	.154	.026	.403	.028	.752	.014	.901	.006
200	ω^2	.000	.003	.050	.017	.151	.026	.401	.028	.751	.014	.900	.006
	R_c^2	.000	.003	.050	.017	.151	.026	.401	.028	.751	.014	.901	.006